Last Thursday, I collected five ideas totally after took part in three groups to discuss the idea of the final project. Some of them look similar. I merged the analogous ideas and got rid of two ideas that focus on low-level design. Now, I am interested at the idea about doing a distributed system for storing data.

The project runs in three steps. Every time when you run the program, the first step is dividing the data to several parts in equal size. The second step is utilizing different machines in Amazon Cloud to store those different files respectively. The third step is to retrieve each file from each machine and merge them when users want to get the specific file he stored before.

Now, I think this project has three difficulties. The first one is that I have no idea about the network knowledge of Cloud Storage. The second one is about the process of merging. The sequence must be correct or the merged file cannot be read correctly. Also, the responsive time should be very fast. Otherwise, users will lost their patience. Thirdly, if the user changes one part of a file, how can we replace the exact part of file in the cloud? Now, my idea is using a revision control to decide which modification is the latest one.

I hope the project should be like a small Dropbox finally.

Yanda Lu